



King's Research Portal

DOI:

[10.1074/jbc.RA118.007347](https://doi.org/10.1074/jbc.RA118.007347)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Kuckelkorn, U., Stübler, S., Textoris-Taube, K., Killian, C., Niewianda, A., Henklein, P., Janek, K., Stumpf, M. P., Mishto, M., & Liepe, J. (2019). Proteolytic dynamics of human 20S thymoproteasome. *Journal of Biological Chemistry*, 294(19), 7740-7754. <https://doi.org/10.1074/jbc.RA118.007347>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Proteolytic dynamics of human 20S thymoproteasome

Ulrike Kuckelkorn¹, Sabine Stübler^{2,3}, Kathrin Textoris-Taube^{4,5}, Christiane Kilian¹, Agathe Niewianda^{4,5}, Petra Henklein¹, Katharina Janek^{4,5}, Michael P.H. Stumpf², Michele Mishto^{4,6,\$,*}, Juliane Liepe^{2,7,\$,*}

- ¹ Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Institut für Biochemie, Germany, 10117 Berlin, Germany
- ² Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London, SW7 2AZ London, UK
- ³ Mathematical Modelling and Systems Biology, Institute of Mathematics, University of Potsdam, 14469 Potsdam, Germany
- ⁴ Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institut für Biochemie, Germany, 10117 Berlin, Germany
- ⁵ Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Shared Facility for Mass Spectrometry, 10117 Berlin, Germany
- ⁶ Centre for Inflammation Biology and Cancer Immunology (CIBCI) & Peter Gorer Department of Immunobiology, School of Immunology & Microbial Science, King's College London, SE1 1UL London, United Kingdom
- ⁷ Max-Planck-Institute for Biophysical Chemistry, 37077 Göttingen, Germany

* Authors equally contributed

\$ **Corresponding authors:**

Michele Mishto; phone: +44 20 078486907; email: michele.mishto@kcl.ac.uk.

Juliane Liepe; phone: +49 5512011471; email: juliane.liepe@mpibpc.mpg.de

Running title: Proteolytic dynamics of human thymoproteasome

Keywords: proteasome, protein degradation, proteolysis, computational biology, antigen processing.

Abstract

An efficient immunosurveillance of CD8⁺ T cells in the periphery depends on positive/negative selection of thymocytes and thus on the dynamics of antigen degradation and epitope production by thymoproteasome and immunoproteasome in the thymus. Although studies in mouse systems have shown how thymoproteasome activity differs from that of immunoproteasome and strongly impacts on the T cell repertoire, the proteolytic dynamics and the regulation of human thymoproteasome are unknown. By combining biochemical and computational modeling approaches, we show here that human 20S thymoproteasome and immunoproteasome differ not only in the proteolytic activity of the catalytic sites but also in the peptide transport. These differences impinge upon the quantity of peptide products rather than where the substrates are cleaved. The comparison of the two human 20S proteasome isoforms depicts different processing of antigens that are associated to tumors and autoimmune diseases.

Introduction

The immune system constantly patrols the human body to detect pathological situations. An important role in this is played by CD8⁺ T cells, which recognize and kill infected and aberrant cells. Because of their high specificity and cytotoxic activity, CD8⁺ T cells are intensely investigated as tools and/or targets of immunotherapies against infection, autoimmunity and cancer (1-3). CD8⁺ T cells recognize, via their T cell receptor (TCR) coupled to a CD8 molecule, a specific epitope presented in the cleft of major histocompatibility complex class I (MHC-I) molecules (4). The cytotoxic CD8⁺ T cells are primed in lymph nodes. The naïve CD8⁺ T cells arrive in the lymph nodes from the thymus. In this latter organ, thymocytes, prior to becoming mature naïve CD8⁺ T cells, undergo a series of maturation/selection processes called central tolerance. According to one of the most accepted models, central tolerance can be summarized in two selection steps. Initially, thymocytes undergo a positive selection step, which takes place in the thymic cortex, leading to the survival and

maturation of double-positive thymocytes that express TCRs with intermediate affinity and/or avidity for MHC-I-peptides complexes. Afterwards, in the thymic medulla thymocytes undergo the negative selection step, which leads to the elimination of thymocytes recognizing self-peptide-MHC complexes with a high affinity (5).

The large majority of peptides bound to MHC-I molecules and recognized by CD8⁺ T lymphocytes are generated by proteasome, which is the final effector of the ubiquitin-proteasome system (6). This barrel shaped protease can break proteins and release the peptide fragments or re-ligate them, thereby forming new (spliced) peptides with sequences that do not recapitulate the parental protein (4). The proteasome is a multi-subunit enzyme, which has a 20S proteasome as the core and various proteins bound at both side of its gate, where they play a regulatory role (4). The 26S proteasome, comprising of a 20S proteasome core coupled to a 19S regulatory complex, is often the most active form of proteasome, with an increasing amount of evidence suggesting that the 20S proteasome is independently functional and both degrades and activates proteins in cells (4,7,8). The 20S proteasome is constituted of four rings, two α rings at the apexes and two β rings forming the central chamber. Each ring has 7 distinct subunits. Each β ring carries three catalytic (i.e., β 1, β 2, and β 5) subunits, which have distinct preferences for peptide sequence motifs (9). Human cells can express different isoforms of catalytic subunits, which are incorporated in distinct proteasome isoforms. Standard proteasome (s-proteasome) contains β 1, β 2, and β 5 subunits. Immunoproteasome (i-proteasome) contains β 1i, β 2i, and β 5i subunits and it is present in immune cells (constitutively) as well in cells exposed to inflammatory milieu. The majority of cells express a mixed-type proteasome population where both s- and i-proteasome subunits are present in various amount (4). A decade ago, Murata and colleagues (10) identified the so-called thymoproteasome (t-proteasome), which carries the β 1i, β 2i, and β 5t subunits and has so far only been detected in the thymus.

To investigate the dynamics and its regulation of peptide hydrolysis by the 20S proteasome and to elucidate differences between s- and i-proteasomes, we previously developed a computational mechanistic model of proteasome peptide degradation (11) (**Fig. 1A**). The model was constructed and compared with many competing models in a Bayesian model selection framework (12,13) and finally challenged with further experimentation.

Specifically, a series of models was constructed with increasing complexity, starting from the simplest Michaelis-Menten model and ignoring the structural properties of the proteasome. Kinetic time-course data tracking the degradation of short fluorogenic peptides by 20S proteasomes was used to test whether the constructed models were able to produce the experimentally observed kinetics. Latter data and further experimentation were used to yield insight into possible dynamics of the proteasome and guide the development of competing mechanistic models with increasing complexity. These models can be seen as representing competing hypotheses, allowing selection of the hypothesis that best justifies our experimental data. In contrast to many hypothesis testing techniques, Bayesian model selection allows us to not only reject a likely wrong hypothesis, but also rank competing models given experimental data (14,15).

Model development and model selection have two aspects to consider: the model structure and the model parameters. While the model structure can be seen as a map, the model parameters essentially identify where in that map the dynamics of the model can occur. It is therefore of interest to define both, model structure and model parameters, or in Bayesian terms the posterior model distribution and the posterior parameter distribution (16,17). Our Bayesian model selection framework allows us to determine both the best model structure and the model parameters, each with prior knowledge, *i.e.* the original model distribution (the construction of a set of competing models and our initial confidence in them) and the original parameter distribution (the allowed values of a kinetic parameter). Both, the suggested model structure and the suggested model parameters, will

contain uncertainty, which can be assessed from the posterior distribution (18,19). Roughly speaking, the broader a posterior parameter distribution is, the less information the experimental data has contained about this parameter and thus the higher the uncertainty. This parameter uncertainty will be carried into model predictions. To note, in some cases not all model parameters must be inferred with low uncertainty to make precise model predictions (20). This is because the systems dynamics may be susceptible to alterations in those parameters.

The applied Bayesian model selection framework allowed us to determine the best model out of ten constructed competing models, however, this does not guarantee the accuracy of the winning model. A crucial step during model development is model validation, whereby independent experimental data, which were not used for model development or model calibration, are used to challenge the chosen model. At this point it is important to note that any model is a simplification of the true biochemical system and that in many cases a model is developed to explain certain aspects of the system, but not all (21). Assumptions and simplifications often dictate under which conditions the system can be described by the developed model. For example, in our 20S proteasome model we do not include proteasome activators such as PA28 $\alpha\beta$ and 19S complexes, which may significantly alter the observed dynamics and its regulation. Model validation can help to elucidate the limits and predictive potential of the model.

In most cases the motivation behind model development is to generate a model to predict system behaviour that cannot be observed experimentally. However, here we have created a model which tests different mechanistic hypotheses and derives kinetic parameters to in turn characterize different proteasome isoforms. As we showed that proteasome isoforms differ quantitatively but not qualitatively in the peptides they generate (22) (at least with the sensitivity allowed by the assays applied), we can therefore, assume that the overall model structure is the same for different isoforms and that only the kinetic parameters differ. These differences can be acquired by comparing

the marginal posterior parameter distributions in a practical rather than statistical manner. That is, differences are detected if the distributions to be compared overlap only slightly and their values with the highest densities clearly vary. On the contrary, if the parameter distributions to be compared cover the same range of possible parameter values then either the data does not contain sufficient information to detect differences, or the parameters indeed do not differ.

In our previous study, parameter inference and subsequent comparison of posterior distributions of the kinetic model parameters (**Fig. 1B**) showed that 20S s- and i-proteasome differ in the activity of their catalytic sites, in peptide transport along their inner channels, and in the transport regulation dynamics. Because the peptide transport seems to often be the rate limiting step of the overall peptide-bond hydrolysis (11), such differences can impinge upon the degradation rate of specific proteins and the generation efficiency of specific antigenic peptides. Quantitative differences in antigenic peptide production, after the downstream antigen presentation steps, can result in an impaired or enhanced CD8⁺ T cell response *in vivo* (22-25).

The efficiency of the antigenic peptide generation by i-proteasome plays a key role in the negative selection of thymocytes since medullary professional antigen presenting cells (APCs) mainly express this proteasome isoform (5). The cortical thymic epithelial cells (cTECs), on the contrary, mainly express the t-proteasome, although s-proteasome catalytic subunits have also been detected (10,26). T-proteasome influences the CD8⁺ T cell repertoire and the response to infection in mice (10,27-29). Some evidence hints towards a unique t-proteasome proteolytic activity, which, in mice, would lead to the generation of t-proteasome-specific antigenic peptides with peculiar features promoting the positive selection of thymocytes (5,28,30).

As a consequence, the difference between the proteolytic dynamics of human i- and t-proteasome is supposed to have a large impact on the central tolerance, and, thus, on the T cell repertoire and on the efficacy of CD8⁺ T cells to recognize

infected or aberrant cells and eliminate them in the human body.

To study in which aspects of proteolytic dynamics the two human proteasome isoforms diverge, we have coupled biochemical experiments to bioinformatics analyses. We have made use of the previously developed computational model in order to infer the kinetic parameters of human 20S t-proteasomes in comparison to human 20S s- and i-proteasomes. We have taken advantage of Bayesian inference to obtain posterior parameter distributions that capture not only the most plausible parameter values, but also the information and uncertainty carried by the experimental data. The latter is of particular interest when aiming to detect differences between the proteasome isoforms. Based on the inferred kinetic parameters, we have performed model simulations to identify the rate limiting steps and peptide transport dynamics of human 20S t-proteasome.

In a second step of the study, we have investigated the quantitative differences between the three 20S proteasome isoforms in substrate cleavage-site preferences and epitope production. Because of the immunological implications that such differences can have, we have used synthetic polypeptides substrates derived from tumors and multiple sclerosis, which are two examples of diseases where the MHC-I-presented epitopes are therapeutically relevant.

Results

Human 20S t-proteasome differs from s- and i-proteasomes in its proteolytic dynamics

Human $\beta 5t$ subunit has been previously detected in different forms of human thymoma (31,32), which are tumors originating from the epithelial cells of the thymus. We have tested whether the t-proteasome subunit was detectable in other cancer-derived cell lines. The mRNA of PSMB11, which is the gene encoding the human $\beta 5t$ proteasome subunit, is also detectable in several tumor-derived or immortalized cell lines by RT-PCR (**Fig. 2A**); however, its expression does not lead to a detectable quantity of $\beta 5t$ subunit through the use of a standard proteomics strategy (**Fig. 2B**). This result

confirms that in humans the expression of the $\beta 5t$ proteasome subunit also seems to be limited to the thymus.

Therefore, not having access to enough human cTECs, we generated a cell line, C5.5. This cell line is derived from the human lymphoblastoid cell line T2. The C5.5 cell line expresses mainly the $\beta 1i$, $\beta 2i$, and $\beta 5t$ subunits, with a $\beta 5:\beta 5t$ subunit ratio of 1:2.5, according to our quantitative proteomics analysis carried out with AQUA peptides (**Fig. 2C,D**). This mixed-type 20S proteasome – here referred to as t-proteasome – purified from the C5.5 cell line has been compared in our study to either the 20S s-proteasome derived from parental T2 or the intermediate-type 20S proteasome purified from EBV-immortalized lymphoblastoid cell lines (LCLs), which has a $\beta 5:\beta 5i$ subunit ratio of 1:2.5 or larger (22), and has often been used as an example of i-proteasome.

To investigate the proteolytic dynamics of human 20S t-proteasome as compared to 20S s- and i-proteasomes, we have adopted an approach that integrates an extensive set of *in vitro* degradation kinetics with computational modeling (**Fig. 1A-B**). We have first purified 20S proteasomes from T2, LCL and C5.5 cells and used these proteasomes to perform *in vitro* degradation kinetics of the short fluorogenic peptides Suc-LLVY-MCA, and Z-LLE-MCA, two substrates specific for the chymotrypsin-like and caspase-like activity of proteasomes. The MCA group of these substrates is released upon endopeptidase cleavage by proteasomes and its fluorescence can be measured quantitatively. Murata *et al.* (10) showed that the degradation rate of Suc-LLVY-MCA – which is mainly carried out by the $\beta 5/\beta 5i/\beta 5t$ subunits – diverges between mouse i- and t-proteasome. In contrast, the cleavage rate of Z-LLE-MCA – which is mainly carried out by the $\beta 1/\beta 1i$ subunits – should not be different between human i- and t-proteasomes, since they carry similar amounts of the $\beta 1i$ subunit (**Fig. 2C**). Furthermore, although these short fluorogenic substrates do not recapitulate the full substrate specificity of proteasomes (22), they have been successfully used to discriminate between 20S s- and i-proteasome dynamics by the development of a computational model and its calibration with time course data (11) (**Fig. 1**). We have here applied the

same computational modeling and model calibration approach on Z-LLE-MCA and Suc-LLVY-MCA digestion kinetics of 20S s-, i- and t-proteasomes. Using Bayesian inference techniques, we have obtained model fits to the experimental data by estimating the model parameters resulting in a posterior parameter distribution for each 20S proteasome isoform (**Fig. S1, S2**). The posterior parameter distribution contains information on possible kinetics parameter values able to explain the experimental datasets and their relationship with each other.

The kinetic differences between 20S s- and i-proteasomes confirm our previous results (11) and therefore the correct set up of our study (**Fig. 3A,B**).

Regarding the comparison of the two proteasome isoforms mainly involved in the positive/negative selection, no difference emerges between 20S t- and i-proteasomes when the estimated parameter distributions obtained from Z-LLE-MCA digestions are compared (**Fig. S3**). On the contrary, in the Suc-LLVY-MCA degradation kinetics we have found differences related to the active-site parameters between these two 20S proteasome isoforms (**Fig. 3A**). Most apparent is the hydrolysis strength k_p (for explanations see **Table 1**), which appears to be approximately 4-fold smaller in 20S t-proteasome compared to 20S i-proteasome and therefore recapitulates the k_p observed in s-proteasome. Also, the dissociation constant of the peptide to the substrate binding site (K_{as}), which is slightly increased in 20S t-proteasome compared to 20S i-proteasome, results again in a similar K_{as} as observed in 20S s-proteasome (**Fig. 3A**). In addition, the peptide transport and transport regulation parameters differ, indicating that the subunit exchange has not only local but also global effects on 20S proteasome dynamics (**Fig. 3B**).

Peptide transport dominates the substrate degradation in human t-proteasome

By studying inferred posterior parameter distribution, we can determine the rate-limiting steps of the reaction. Previously, we showed that for both 20S s- and i-proteasomes, the gate conformation, which determines the peptide transport (influx and efflux), is often the rate-limiting step (11). By *in silico* simulations

we have now found that although several kinetics parameters significantly differ between 20S i- and t-proteasome, for the Suc-LLVY-MCA substrate the rate-limiting step of 20S t- and i-proteasomes is the peptide transport (**Fig. 3C**), whereas for the Z-LLE-MCA substrate the rate-limiting steps are primarily the peptide-bond hydrolysis at the active site and secondarily the peptide transport (**Fig. 3D**).

Peptide transport also regulates how much substrate and product are located inside the proteasome chamber over time (11). The local substrate concentration around the active site Thr₁ inside the proteasome chamber then strongly influences peptide hydrolysis. For a fast substrate turnover, the peptide flux through the chamber should be large enough to allow sufficient supply of new substrate molecules and sufficient efflux of product molecules. Furthermore, the peptide flux should be so that the substrate concentration around the active site is high enough to obtain reaction velocities at a level of v_{max} , but low enough to avoid substrate inhibition (11). Taking this into account, differences in peptide transport between proteasome isoforms, and thus in the filling dynamics of the proteasome, should strongly influence the observed substrate degradation rates. To investigate this aspect, we have used the estimated posterior parameter distributions for 20S i- and t-proteasomes to calculate *in silico* the amount of substrate and products inside the proteasome chamber over time. In the case of Z-LLE-MCA, our simulations show only minor differences in the filling kinetics between 20S i- and t-proteasomes (**Fig. 4A**). In contrast, in the case of Suc-LLVY-MCA, our simulations suggest that the chamber of 20S t-proteasome is filled more slowly with substrate and product molecules than the 20S i-proteasome chamber (**Fig. 4B**). For both 20S proteasome isoforms, our simulations suggest that, in our experimental conditions, after 6 h of reaction an equal proportion of substrate and product molecules are present inside the chamber. However, this equal proportion is reached faster by 20S i-proteasome than 20S t-proteasome, indicating a stronger peptide bond hydrolysis activity by 20S i-proteasome, which could be reflected by its higher k_p

value (**Fig. 3A**). On the contrary, the peptide-bond hydrolysis of the Z-LLE-MCA substrate does not differ between these two isoforms and it is much lower compared to that of the Suc-LLVY-MCA substrate.

In summary, these analyses highlight that the overall substrate degradation differences between human 20S i- and t-proteasome cannot be explained only by differences in the active site subunits, but they can result from dynamical differences that regulate the peptide transport efficiencies.

20S proteasome isoforms quantitatively differ in substrate cleavage-site strength and the generation of self epitopes

To introduce a further degree of complexity in our experimental approach, we have analyzed, by mass spectrometry (MS), digests obtained after incubation of 20S proteasomes with eight synthetic polypeptides - three derived from the melanoma-associated antigen gp100^{PMEL17}, and five from myelin sheath proteins, which are the main auto-antigens attacked by CD8⁺ T cells in multiple sclerosis (**Table S1**). We have selected these substrates because of their immunogenicity toward CD8⁺ T cells associated with either tumor recognition or autoimmune response against oligodendrocytes in multiple sclerosis. In other words, epitopes derived from these antigenic sequences are known to be detected at the cell surface by autoreactive CD8⁺ T cells. These lymphocytes specifically detect tumor-associated antigens - e.g. gp100 - and multiple sclerosis-associated antigens - e.g. myelin oligodendrocyte glycoprotein (MOG) and myelin basic protein (MBP) - which can be expressed in medullary thymic APCs. Nonetheless, they survived negative selection in the thymus and are present in the periphery as shown in various studies (33-37).

By MS analysis of the peptide products generated by 20S s-, i- and t-proteasomes we have identified 510 non-spliced and 67 spliced peptide products. All spliced and non-spliced peptide products are quantifiable in the digestions. All 20S proteasome isoforms cleave the substrates between the same residues.

The fact that any of the three 20S proteasome isoforms do not use even one

substrate cleavage site that is not used by the other two proteasome isoforms does not exclude that substrate cleavage sites can be preferentially used by one of the 20S proteasome isoforms. We have verified this hypothesis through investigation of proteasome-mediated digestion kinetics of four substrates (*i.e.* gp100₂₀₁₋₂₃₀, gp100₃₅₋₅₇, MOG₁₇₂₋₂₀₂, and MBP₁₀₂₋₁₂₉). We have adopted a quantitative strategy in our MS analysis, using the Quantification with Minimal Effort (QME) methodology. QME estimates the absolute content of spliced and non-spliced peptide products based on their MS peak area, measured in the digestion probe (38). QME can also estimate the frequency of use of each substrate cleavage-site by proteasomes, *i.e.* in the substrate cleavage-site strength (SCS).

According to our hypothesis, we have observed quantitative differences in the substrate cleavage-site predominantly used, between the three 20S proteasome isoforms (**Fig. 5**). There are no amino acids that are clearly preferred by one of the 20S proteasome isoforms rather than the others for peptide-bond hydrolysis, likely because the peptide sequence motifs (8-10 residues) surrounding the cleavage site influence the frequency of usage of that cleavage site. For instance, the gp100 L₂₂₅ residue is seldom used by 20S t-proteasome (**Fig. 5A**), although the MBP L₁₁₂ (**Fig. 5B**) and the MOG L₁₉₃ residues are scarcely used by 20S i-proteasome (**Fig. 5C**), which conversely prefers the gp100 L₃₉ residue (**Fig. 5D**). For a more systematic comparison of the SCSs of the 20S proteasome isoforms we performed pairwise correlations between their SCSs. We have found significant pairwise correlations between SCSs of all three 20S proteasome isoforms, thereby indicating that their overall catalytic activity is comparable (**Fig. 6**). However, when we analyzed the SCS of specific substrates, differences emerge. These differences are due to specific substrate cleavage-sites, which are used by all three 20S proteasome isoforms although with divergent frequencies (*e.g.*, gp100 F₂₁₅, L₂₂₅, D₂₂₆, L₃₉, and A₅₅).

These quantitative differences in SCSs are also reflected in the generation of specific peptide products, including some that have already been shown to be epitopes recognized by CD8⁺ T cells. For instance,

the epitope MBP₁₁₁₋₁₁₉, which is a HLA-A*02:01-binding epitope recognized by multiple sclerosis and healthy donor patients (39-43), is better generated by 20S t-proteasome than 20S i-proteasome. However, this phenomenon is epitope-specific since the generation of the epitope MBP₁₀₇₋₁₁₅ (41) is not favored when carried out by human 20S t-proteasome as compared to 20S i-proteasome (**Fig. 7**).

Discussion

Diverging from what was shown for mouse 20S t-proteasome (30), human 20S t-proteasome does not seem to possess a unique proteolytic activity in processing the self-antigen substrates included in this study and in our experimental conditions. We show here, however, that human 20S t- and i-proteasome differ in their catalytic activity, peptide transport and transport regulation. The differences in the peptide transport are particularly relevant since the latter is often the rate-limiting step, at least in the degradation of short peptides, as also shown here for human 20S t-proteasome. Differences in the peptide transport imply alterations of the overall 20S proteasome dynamics and long-range effects over the entire proteasome chamber due to the incorporation of the $\beta 5t$ subunit. As a consequence, the quantity of peptides produced by the $\beta 1i$ and $\beta 2i$ subunits (present in both t- and i-proteasome) can differ in case they are incorporated into either 20S t- or i-proteasome, since the peptide transport, and therefore the concentration and dynamics of peptides in these proteolytic pockets, could be altered too.

These dissimilarities in catalytic activity, peptide transport and transport regulation can explain the variable preferences of human proteasome isoforms for specific substrate-cleavage sites. The fact that we have not observed cleavage sites only used by either 20S t- or i-proteasome could not be explained by the presence of standard catalytic subunits in our 20S t- and i-proteasome preparations. Indeed, if a substrate cleavage site is used only by 20S i-proteasome, for instance, we shall not detect it in the 20S t-proteasome digestions and vice versa. Therefore,

although in our experimental set up we cannot exclude the existence of substrate cleavage sites exclusively used by human 20S s-proteasome, we can exclude substrate cleavage sites exclusively used by either 20S i- or t-proteasome in the sequences analyzed in this study, as previously demonstrated already for human and mouse 20S i-proteasomes (22). This does not exclude the possibility that improving MS sensitivity could show a qualitative difference in the spliced and non-spliced products between 20S proteasome isoforms and this also does not provide direct information about differences between 20S proteasome isoforms coupled to 19S and other regulatory complexes.

We have observed, however, strong variation from proteasome isoform to proteasome isoform in the quantity of peptides produced, including MHC-I-binding epitopes, as shown here for the epitope MBP₁₁₁₋₁₁₉. These quantitative differences in peptide generation can lead to a negligible presentation of epitopes at the cell surface as demonstrated in previous studies where mouse 20S s- and i-proteasomes have been compared (22,24,25). Our observations, therefore, are compatible with the hypothesis that cells expressing human

thymoproteasome (both 20S proteasomes and 20S proteasomes coupled to regulatory complexes) could present on the cell surface some specific "private" epitopes, which could be directly involved in the positive selection of thymocytes (5). These "private" epitopes were not presented by cells expressing i-proteasome - although they were produced in the intracellular space - because of the progressive reduction of the peptide amounts during the steps of the antigen presentation pathway (44).

Different proteolytic dynamics between human i- and t-proteasomes could also result in significant variation in the turnover of specific antigens, and therefore in the antigenic landscape of cells expressing either t- or i-proteasome, which would further favor the "private epitopes" hypothesis.

Our study also shows that 20S t-proteasomes can generate spliced peptides, which are thought to represent a large portion of the MHC-I immunopeptidome; the peptides bound to MHC-I molecules (45-47). The generation of spliced epitopes by t- and i-proteasomes in the thymus could strongly impinge upon our models of central tolerance and discrimination between self and non-self by our immune system (48).

Material & Methods

Cell lines

To generate a cell line stably overexpressing the proteasome $\beta 5t$ subunit (protein: A5LHX3; gene: PSMB11), the thymic cDNA was transcribed from human thymic total RNA (Clontech Laboratories) by the Transcriptor First Strand cDNA Synthesis Kit (Roche Life Science) according to the manufacture's instructions. For amplification of the human $\beta 5t$ sequence (PSMB11) by PCR the following primers was used: fw 5'gggatggctctgcaggatgtgtgc and rev 5'ctcacaccgtctcagtcctgc. The PCR product was first inserted into pcDNA3.1(+) (Thermo Fisher Scientific) and then re-cloned into a pSG5 vector (Stratagene) via EcoR1/BamH1. T2 cells were cultured in RPMI supplemented with fetal bovine serum to a final concentration of 10% FBS, and 2 mM L- glutamine in 5% CO₂ atmosphere. Cells were transfected with pSG5/PSMB11, pSG5/PSMB9 (Lab stock) and pSVneo in

equal amounts using the Amaxa Cell line Nucleofector® Kit C (Lonza). Stable transfected cells were selected with 1 mg/ml G418 and the expression was controlled by PCR and western blotting. The positive cells were isolated and cultured again. The clone C5.5, which was transfected with $\beta 5t$ (PSMB11) and $\beta 1i$ (PSMB9) subunits, was selected for the study.

To verify the endogenous expression of the $\beta 5t$ subunit the following cell lines were grown in basal ISCOVE media supplemented with 10% FCS, 2mM glutamine and PenStrep in a 5% CO₂ atmosphere: i. T2 and C5.5 cell lines; ii. LCLs; iii. HeLa cell, 33.2 and 21.2 clones; the latter are stably transfected with HLA-A*02:01 and $\beta 1i+\beta 2i+\beta 5i$ subunits or $\beta 1i+\beta 2i$ subunits respectively (49); iv. DU145 prostate carcinoma purchased from ATCC (ATCC® HTB-81™); v. RKO human colon carcinoma (ATCC CRL-2577); vi. MeWo malignant melanoma (ATCC® HTB-65™); vii. human umbilical vein endothelial cells (HUVEC); viii. THP1

acute monocytic leukemia (ATCC® TIB-202™); ix. Mel15 malignant melanoma (50); x. SaOs human bone osteosarcoma (ATCC HTB-85); A549 adenocarcinomic human alveolar basal epithelial cells (ATCC® CCL-185™).

Western blots and $\beta 5t$ subunit identification and quantification

Proteasome subunits have been revealed by western blot assays as previously described (22) using: anti- $\alpha 4$ (1:10000; in-house), anti- $\beta 1$ (1:1000; Enzo Life Science), anti- $\beta 2$ (1:1000; Enzo Life Science), anti- $\beta 5$ (1:1000; Abcam), anti- $\beta 1i$ (1:20000; in-house), anti- $\beta 2i$ (1:3000; Thermo Fisher), anti- $\beta 5i$ (1:1000; Santa Cruz Biotechnology), and anti-*flag* (1:1000; Sigma-Aldrich).

For the identification of $\beta 5t$ subunit in different cell lines (**Fig. 2B**), bands have been excised from SDS-PAGE gel, washed with 200 mM ammonium bicarbonate in acetonitrile/water (50:50, v/v) and 50 mM ammonium bicarbonate in acetonitrile/water (5:95, v/v) (digestion buffer), shrunk by dehydration in acetonitrile and dried. For reduction of cysteine residues 50 μ l of 10 mM DTT in 100 mM ammonium bicarbonate have been added and the samples have been incubated for 45 mins at 56°C. After 3 times washing with water, 50 μ l 55 mM iodoacetamide in 50 mM ammonium bicarbonate have been added. Incubation has been performed at room temperature in the dark for 20 mins. The gel bands have been washed with water and digestion buffer, and dried in acetonitrile after shrinking. 25 μ l 0.1 g/l trypsin (Promega) in digestion buffer has been added and incubated at 37°C overnight. Tryptic-digested peptides have been extracted by one change of 25 μ l 0.5% TFA in acetonitrile/water (70:30, v/v), and 25 μ l acetonitrile (10 min for each) at room temperature, combined and dried down.

LC-MS/MS analyses of peptides have been performed as following: the sample has been concentrated for 4 mins on a trap column (PepMap C18, 5 mm x 300 μ m x 5 μ m, 100Å, ThermoFisher Scientific) with 2:98 (v/v) acetonitrile/water containing 0.1% (v/v) trifluoroacetic acid at a flow rate of 30 μ l/min and then analyzed by nanoscale LC-MS/MS measurements using a Q Exactive Plus mass spectrometer coupled with an Ultimate 3000 RSLCnano (ThermoFisher

Scientific). The system comprised a 75 μ m i.d. x 250 mm nano LC column (Acclaim PepMap C18, 2 μ m; 100 Å; ThermoFisher Scientific). The mobile phase (A) was 0.1% (v/v) formic acid (FA) in water, and (B) was 80:20 (v/v) acetonitrile/water containing 0.1% (v/v) FA. The elution has been carried out using a gradient 3-43% B in 80 min with a flow rate of 300 nl/min. Full MS spectra (m/z 350–1,600) have been acquired at a resolution of 70,000 (FWHM) followed by a data-dependent MS/MS fragmentation of the top10 precursor ions (resolution 17,500, 1⁺ charge state excluded, isolation window of 1.6 m/z , normalized collision energy of 27%, dissociation method HCD). The maximum ion injection time for MS scans has been set to 50 ms and for MS/MS scans to 120 ms. Background ions at m/z 391.2843 and 445.1200 act as lock mass. Protein identifications has been performed with Mascot software version 2.6.1. (Matrix Science Ltd., London, UK). Data have been searched against SwissProt (July 2017), taxonomy human (20,215 sequences) and a contaminant database (247 sequences). The following parameters have been set: enzyme: trypsin/P with one missed cleavage, static modification: carbamidomethylation (C), variable modifications: protein N-acetylation and oxidation (M), mass tolerances for MS and MS/MS: 5 ppm and 0.02 Da. Proteins have been accepted as identified if at least two unique peptides provided a Mascot MS/MS score for identity ($p < 0.01$).

For relative protein quantification of $\beta 5$ and $\beta 5t$ subunits, two heavy AQUA peptides (**Table S2**) have been synthesized based on Fmoc solid phase chemistry, i.e. the heavy analogue peptide $^{226}\text{DAYSGGAVNLYHVR}_{239}$ (with $[M+H]^+_{\text{exp}} = 1528.80$) of human $\beta 5$ subunit and the heavy analogue peptide $^{216}\text{DAYSGGSVDLFHVR}_{229}$ (with $[M+H]^+_{\text{exp}} = 1529.79$) of the human $\beta 5t$ subunit. During peptide synthesis, the heavy isotope-labeled amino acid Leucine-N-Fmoc ($U-^{13}\text{C}_6$, N^{15}) from Euriso-Top GmbH (Saarbrücken, Germany) has been incorporated. Estimation of exact peptide amount has been performed by amino acid analysis (Gennaxon Bioscience, Ulm Germany).

$\beta 5/\beta 5t$ subunit quantification of purified C5-5 20S proteasome reported in **Fig. 2D**

has been performed after LC separation of tryptic peptides on a MALDI-TOF/TOF mass spectrometer. In the assay ($n = 2$), 5 μg proteasome have been reduced, alkylated and digested with trypsin as described elsewhere (51). Aliquots of the sample spiked with 500 fmol of each AQUA peptide have been analyzed by LC-MS/MS on a 4700 Proteomics Analyzer (ABSCIEX, Framingham, MS) off-line coupled with a Dionex UltiMate 3000 RSLC system and Probot fractionation device (Thermo Scientific, Idstein, Germany) as previously described (52). The calculation of the absolute amount of the tryptic peptides has been performed by comparison of the MS peak areas with those of the corresponding AQUA peptides. Based on the absolute amounts, the relative ratio $\beta 5:\beta 5t$ has been determined as follows: $\beta 5t/(\beta 5+\beta 5t)$.

RT-PCR and validation

Total RNA has been isolated from human cell lines by High Pure RNA Isolation Kit (Roche) in presence of DNase according manufactures protocol. The cDNA has been obtained from 1 μg total RNA by Primer Script RT Reagent Kit (TaKaRa Bio Inc.). The RT-PCR has been performed with 1 μl cDNAs for 25 cycles at an annealing temperature of 56°C. For the RT-PCR specific for the huPSMB11 gene, the following primers have been used: huPSMB11 Fw 5' `gggatggctctgcaggatgtgtgc`, and huPSMB11 Rev 5' `ctcacaccgtctcagtcctctgc`, thereby generating a transcript of 907 bp. To control the efficiency of RT-PCR, actin has been amplified from the cDNA by RT-PCR using the following actin specific primers: actin Fw 5' `ctcacatggatgatatcg` and actin Rev 5' `tcgtcatactctgcttgctg`. For the verification of the $\beta 5t$ -subunit's RT-PCR products the PCR products have been eluted from the agarose gel and inserted into pCR2.1 Topo® (ThermoFisher) and sequenced by using the T7 primer by LCG Genomics GmbH (LGC group). The latter step has confirmed the specificity of the amplified sequence marked in **Fig. 2A**.

Computational analysis of the proteasome dynamics

20S proteasome degradation dynamics have been assessed using an integrative modeling approach with Bayesian model calibration to the *in vitro* degradation of the short fluorogenic peptides Suc-LLVY-MCA, and Z-LLE-MCA. *In vitro*

degradation kinetics ($n=3-5$) have been performed with different substrate concentration (0-480 μM) in 100 μl TEAD buffer (Tris 20 mM, EDTA 1 mM, NaN_3 1 mM, DTT 1 mM, pH 7.2) at 37°C as previously described (22). For the analysis we have used the computational model of 20S proteasome activity previously published (11). In an essence, by applying approximate Bayesian computation to fit the model to the experimental data, we have obtained posterior parameter distributions for each proteasome isoform. The comparison of the marginal posterior parameter distributions has allowed us to detect differences in the kinetics parameters between the different isoforms. For the Bayesian inference we have used the package ABC-SysBio (12) implemented in Python with GPU support (53). The prior distributions for all parameters was uniform, as described in (11). Furthermore, all other algorithm parameters have been kept as in Liepe *et al.* (11). The posterior analysis has been carried out in R (54). Computation of rate-limiting steps has been performed as described in Liepe *et al.* (11).

20S proteasome purification

20S proteasomes have been purified from T2, C5.5 and LCL, as previously described (55). Proteasome concentration has been measured by Bradford staining and verified by Coomassie staining in a SDS-Page gel, as show elsewhere (56). The purity of the standardized proteasome preparations has been previously shown (22). LCL and C5.5 cell lines mainly express human i-proteasome (22,57) or t-proteasomes, respectively (**Fig. 2**). LCL proteasome has been often used as an example of i-proteasome in several previous studies (22,38,39,50,57-61). T2 cell line expresses only s-proteasome (**Fig. 2**). The three proteasomes have been purified in parallel to minimize artifacts due to the purification or storage conditions.

Peptides and *in vitro* digestion of synthetic polypeptides

The sequence enumeration for the polypeptide substrates is reported in **Table S1**. All peptides have been synthesized using Fmoc solid phase chemistry. Synthetic polypeptides (20-40 μM) have been digested by 1-3 μg 20S proteasomes in 100 μl TEAD buffer over

time at 37°C, as previously described (22).

Quantitative analysis of peptide products by QME and MS

Liquid-chromatography MS analyses of polypeptide digestion products have been performed as previously described (52) with the ESI-ion trap instrument DECA XP MAX (ThermoFisher Scientific, USA). Database searching has been performed using the SpliceMet's ProteaJ (52). Quantification of peptides produced in the experiments have been carried out by applying QME method to the LC-MS analyses. QME estimates the absolute content of spliced and non-spliced peptide products based on their MS peak area measured in the digestion probe. QME is an optimization tool that makes use of the law of mass conservation and mass spectrometry instrument features. The QME algorithm parameters were empirically computed in our previous study (38) and have been here applied. By applying QME we have also calculated the SCS, which describes the relative frequencies of proteasome cleavage after any given residue of the synthetic polypeptide substrate. SCS values shown

in this study are the average of SCS measured in kinetics assays.

Statistical analysis

The SCS correlation has been done using Pearson's correlation coefficient. Significance tests have been performed testing for association between paired samples based on Pearson's correlation coefficient. All data analysis have been implemented in R (54). Sample size and number of replicates are disclosed in the figure and table captions.

Data and Software Availability

Datasets are available at the online archive Mendeley (<http://dx.doi.org/10.17632/zjhhc2bm9p.1>).

Datasets linked to the computational model of proteasome dynamics and published by Liepe *et al.* (11), are available at: <http://datadryad.org/resource/doi:10.5061/dryad.nk453>.

The statistical background behind the model inference presented in this study is described in (12,13) and its implementation can be found at ABC-SysBio: <https://sourceforge.net/projects/abc-sysbio/>.

Acknowledgments

We thank C. Keller, P. Kunert, B. Brecht-Jachan for technical assistance (Charité Berlin) and W. Scott and A. Abbas (KCL) for proofreading the text. We thank the Shared Facility Mass Spectrometry of the Charité for support in acquisition (i.a. and analysis) of the data and A. Goede for the use of QME software. This work was financed in part by Cancer Research UK King's Health Partners Centre at King's College London (Development Fund 2018) to MM, by Berlin Institute of Health (BIH, CRG1-TP1) to P. M. Kloetzel, by the BBSRC to MPHS.

The authors have no conflicts of interest.

Author contribution: MM, UK, JL developed the whole project, supervised the data analysis and generation and wrote the manuscript. KTT, KJ, MPHS developed part of the project, performed part of the analyses and/or experiments and wrote part of the manuscript. SS, CK, AN performed analysis and/or experiments. PH provided materials.

References

1. Pearson, H., Daouda, T., Granados, D. P., Durette, C., Bonneil, E., Courcelles, M., Rodenbrock, A., Laverdure, J. P., Cote, C., Mader, S., Lemieux, S., Thibault, P., and Perreault, C. (2016) MHC class I-associated peptides derive from selective regions of the human genome. *J Clin Invest* **126**, 4690-4701
2. Gubin, M. M., Zhang, X., Schuster, H., Caron, E., Ward, J. P., Noguchi, T., Ivanova, Y., Hundal, J., Arthur, C. D., Krebber, W. J., Mulder, G. E., Toebes, M., Vesely, M. D., Lam, S. S., Korman, A. J., Allison, J. P., Freeman, G. J., Sharpe, A. H., Pearce, E. L., Schumacher, T. N., Aebersold, R., Rammensee, H. G., Melief, C. J., Mardis, E. R.,

- Gillanders, W. E., Artyomov, M. N., and Schreiber, R. D. (2014) Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577-581
3. Jaravine, V., Raffegerst, S., Schendel, D. J., and Frishman, D. (2017) Assessment of cancer and virus antigens for cross-reactivity in human tissues. *Bioinformatics* **33**, 104-111
4. Mishto, M., and Liepe, J. (2017) Post-Translational Peptide Splicing and T Cell Responses. *Trends Immunol* **38**, 904-915
5. Klein, L., Kyewski, B., Allen, P. M., and Hogquist, K. A. (2014) Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol* **14**, 377-391
6. Weissman, A. M., Shabek, N., and Ciechanover, A. (2011) The predator becomes the prey: regulating the ubiquitin system by ubiquitylation and degradation. *Nat Rev Mol Cell Biol* **12**, 605-620
7. Solomon, H., Brauning, B., Fainer, I., Ben-Nissan, G., Rabani, S., Goldfinger, N., Moscovitz, O., Shakked, Z., Rotter, V., and Sharon, M. (2017) Post-translational regulation of p53 function through 20S proteasome-mediated cleavage. *Cell Death Differ* **24**, 2187-2198
8. Olshina, M. A., Ben-Nissan, G., and Sharon, M. (2018) Functional regulation of proteins by 20S proteasome proteolytic processing. *Cell Cycle* **17**, 393-394
9. Schrader, J., Henneberg, F., Mata, R. A., Tittmann, K., Schneider, T. R., Stark, H., Bourenkov, G., and Chari, A. (2016) The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. *Science* **353**, 594-598
10. Murata, S., Sasaki, K., Kishimoto, T., Niwa, S., Hayashi, H., Takahama, Y., and Tanaka, K. (2007) Regulation of CD8⁺ T cell development by thymus-specific proteasomes. *Science* **316**, 1349-1353
11. Liepe, J., Holzthutter, H. G., Bellavista, E., Kloetzel, P. M., Stumpf, M. P., and Mishto, M. (2015) Quantitative time-resolved analysis reveals intricate, differential regulation of standard- and immuno-proteasomes. *Elife* **4**
12. Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C. P., and Stumpf, M. P. (2014) A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nat Protoc* **9**, 439-456
13. Liepe, J., Barnes, C., Cule, E., Erguler, K., Kirk, P., Toni, T., and Stumpf, M. P. (2010) ABC-SysBio--approximate Bayesian computation in Python with GPU support. *Bioinformatics* **26**, 1797-1799
14. Vyshemirsky, V., and Girolami, M. A. (2008) Bayesian ranking of biochemical system models. *Bioinformatics* **24**, 833-839
15. Toni, T., and Stumpf, M. P. (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* **26**, 104-110
16. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* **6**, 187-202
17. Putter, H., Heisterkamp, S. H., Lange, J. M., and de Wolf, F. (2002) A Bayesian approach to parameter estimation in HIV dynamical models. *Stat Med* **21**, 2199-2214
18. Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* **3**, 1871-1878
19. Erguler, K., and Stumpf, M. P. (2011) Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Mol Biosyst* **7**, 1593-1602
20. Liepe, J., Filippi, S., Komorowski, M., and Stumpf, M. P. (2013) Maximizing the information content of experiments in systems biology. *PLoS Comput Biol* **9**, e1002888

21. Nurse, P., and Hayles, J. (2011) The cell in an era of systems biology. *Cell* **144**, 850-854
22. Mishto, M., Liepe, J., Textoris-Taube, K., Keller, C., Henklein, P., Weberruss, M., Dahlmann, B., Enenkel, C., Voigt, A., Kuckelkorn, U., Stumpf, M. P., and Kloetzel, P. M. (2014) Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation. *Eur J Immunol* **44**, 3508-3521
23. Zanker, D., Waithman, J., Yewdell, J. W., and Chen, W. (2013) Mixed Proteasomes Function To Increase Viral Peptide Diversity and Broaden Antiviral CD8+ T Cell Responses. *J Immunol* **191**, 52-59
24. Kincaid, E. Z., Che, J. W., York, I., Escobar, H., Reyes-Vargas, E., Delgado, J. C., Welsh, R. M., Karow, M. L., Murphy, A. J., Valenzuela, D. M., Yancopoulos, G. D., and Rock, K. L. (2011) Mice completely lacking immunoproteasomes show major changes in antigen presentation. *Nat Immunol* **13**, 129-135
25. Deol, P., Zaiss, D. M., Monaco, J. J., and Sijts, A. J. (2007) Rates of processing determine the immunogenicity of immunoproteasome-generated epitopes. *J Immunol* **178**, 7557-7562
26. Nil, A., Firat, E., Sobek, V., Eichmann, K., and Niedermann, G. (2004) Expression of housekeeping and immunoproteasome subunit genes is differentially regulated in positively and negatively selecting thymic stroma subsets. *Eur J Immunol* **34**, 2681-2689
27. Takada, K., Van Laethem, F., Xing, Y., Akane, K., Suzuki, H., Murata, S., Tanaka, K., Jameson, S. C., Singer, A., and Takahama, Y. (2015) TCR affinity for thymoproteasome-dependent positively selecting peptides conditions antigen responsiveness in CD8(+) T cells. *Nat Immunol* **16**, 1069-1076
28. Nitta, T., Murata, S., Sasaki, K., Fujii, H., Ripen, A. M., Ishimaru, N., Koyasu, S., Tanaka, K., and Takahama, Y. (2010) Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells. *Immunity* **32**, 29-40
29. Kincaid, E. Z., Murata, S., Tanaka, K., and Rock, K. L. (2016) Specialized proteasome subunits have an essential role in the thymic selection of CD8(+) T cells. *Nat Immunol* **17**, 938-945
30. Sasaki, K., Takada, K., Ohte, Y., Kondo, H., Sorimachi, H., Tanaka, K., Takahama, Y., and Murata, S. (2015) Thymoproteasomes produce unique peptide motifs for positive selection of CD8(+) T cells. *Nat Commun* **6**, 7484
31. Yamada, Y., Tomaru, U., Ishizu, A., Kiuchi, T., Marukawa, K., Matsuno, Y., and Kasahara, M. (2011) Expression of proteasome subunit beta5t in thymic epithelial tumors. *Am J Surg Pathol* **35**, 1296-1304
32. Yamada, Y., Tomaru, U., Ishizu, A., Kiuchi, T., Kasahara, M., and Matsuno, Y. (2014) Expression of thymoproteasome subunit beta5t in type AB thymoma. *J Clin Pathol* **67**, 276-278
33. Takase, H., Yu, C. R., Mahdi, R. M., Douek, D. C., Dirusso, G. B., Midgley, F. M., Dogra, R., Allende, G., Rosenkranz, E., Pugliese, A., Egwuagu, C. E., and Gery, I. (2005) Thymic expression of peripheral tissue antigens in humans: a remarkable variability among individuals. *Int Immunol* **17**, 1131-1140
34. Adamopoulou, E., Tenzer, S., Hillen, N., Klug, P., Rota, I. A., Tietz, S., Gebhardt, M., Stevanovic, S., Schild, H., Tolosa, E., Melms, A., and Stoeckle, C. (2013) Exploring the MHC-peptide matrix of central tolerance in the human thymus. *Nat Commun* **4**, 2039
35. Perchellet, A., Stromnes, I., Pang, J. M., and Gorman, J. (2004) CD8+ T cells maintain tolerance to myelin basic protein by 'epitope theft'. *Nat Immunol* **5**, 606-614
36. Perchellet, A., Brabb, T., and Gorman, J. M. (2008) Crosspresentation by nonhematopoietic and direct presentation by hematopoietic cells induce central tolerance to myelin basic protein. *Proc Natl Acad Sci U S A* **105**, 14040-14045

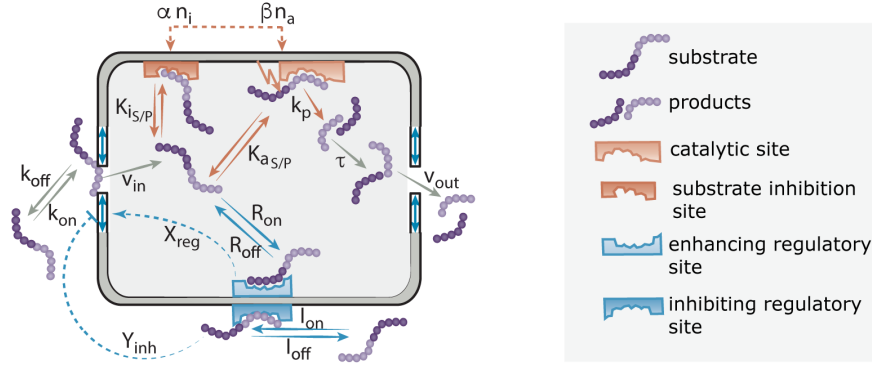
37. Friese, M. A., Jakobsen, K. B., Friis, L., Etzensperger, R., Craner, M. J., McMahon, R. M., Jensen, L. T., Huygelen, V., Jones, E. Y., Bell, J. I., and Fugger, L. (2008) Opposing effects of HLA class I molecules in tuning autoreactive CD8⁺ T cells in multiple sclerosis. *Nat Med* **14**, 1227-1235
38. Mishto, M., Goede, A., Taube, K. T., Keller, C., Janek, K., Henklein, P., Niewianda, A., Kloss, A., Gohlke, S., Dahlmann, B., Enenkel, C., and Michael Kloetzel, P. (2012) Driving Forces of Proteasome-catalyzed Peptide Splicing in Yeast and Humans. *Mol Cell Proteomics* **11**, 1008-1023
39. Mishto, M., Bellavista, E., Ligorio, C., Textoris-Taube, K., Santoro, A., Giordano, M., D'Alfonso, S., Listi, F., Nacmias, B., Cellini, E., Leone, M., Grimaldi, L. M., Fenoglio, C., Esposito, F., Martinelli-Boneschi, F., Galimberti, D., Scarpini, E., Seifert, U., Amato, M. P., Caruso, C., Foschini, M. P., Kloetzel, P. M., and Franceschi, C. (2010) Immunoproteasome LMP2 60HH variant alters MBP epitope generation and reduces the risk to develop multiple sclerosis in Italian female population. *PLoS One* **5**, e9287
40. Jurewicz, A., Biddison, W. E., and Antel, J. P. (1998) MHC class I-restricted lysis of human oligodendrocytes by myelin basic protein peptide-specific CD8 T lymphocytes. *J Immunol* **160**, 3056-3059
41. Berthelot, L., Laplaud, D. A., Pettre, S., Ballet, C., Michel, L., Hillion, S., Braudeau, C., Connan, F., Lefrere, F., Wiertlewski, S., Guillet, J. G., Brouard, S., Choppin, J., and Souliou, J. P. (2008) Blood CD8⁺ T cell responses against myelin determinants in multiple sclerosis and healthy individuals. *Eur J Immunol* **38**, 1889-1899
42. Zang, Y. C., Li, S., Rivera, V. M., Hong, J., Robinson, R. R., Breitbach, W. T., Killian, J., and Zhang, J. Z. (2004) Increased CD8⁺ cytotoxic T cell responses to myelin basic protein in multiple sclerosis. *J Immunol* **172**, 5120-5127
43. Tsuchida, T., Parker, K. C., Turner, R. V., McFarland, H. F., Coligan, J. E., and Biddison, W. E. (1994) Autoreactive CD8⁺ T-cell responses to human myelin protein-derived peptides. *Proc Natl Acad Sci U S A* **91**, 10859-10863
44. Princiotta, M. F., Finzi, D., Qian, S. B., Gibbs, J., Schuchmann, S., Buttgerit, F., Bennink, J. R., and Yewdell, J. W. (2003) Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* **18**, 343-354
45. Liepe, J., Marino, F., Sidney, J., Jeko, A., Bunting, D. E., Sette, A., Kloetzel, P. M., Stumpf, M. P., Heck, A. J., and Mishto, M. (2016) A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* **354**, 354-358
46. Faridi, P., Li, C., Ramarathnam, S. H., Vivian, J. P., Illing, P. T., Mifsud, N. A., Ayala, R., Song, J., Gearing, L. J., Hertzog, P. J., Ternette, N., Rossjohn, J., Croft, N. P., and Purcell, A. W. (2018) A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci Immunol* **3**
47. Liepe, J., Sidney, J., Lorenz, F. K. M., Sette, A., and Mishto, M. (2019) Mapping the MHC Class I-Spliced Immunopeptidome of Cancer Cells. *Cancer Immunol Res* **7**, 62-76
48. Liepe, J., Ovaa, H., and Mishto, M. (2018) Why do proteases mess up with antigen presentation by re-shuffling antigen sequences? *Curr Opin Immunol* **52**, 81-86
49. Keller, M., Ebstein, F., Burger, E., Textoris-Taube, K., Gorny, X., Urban, S., Zhao, F., Dannenberg, T., Sucker, A., Keller, C., Saveanu, L., Kruger, E., Rothkotter, H. J., Dahlmann, B., Henklein, P., Voigt, A., Kuckelkorn, U., Paschen, A., Kloetzel, P. M., and Seifert, U. (2015) The proteasome immunosubunits, PA28 and ER-aminopeptidase 1 protect melanoma cells from efficient MART-126-35 -specific T-cell recognition. *Eur J Immunol* **45**, 3257-3268
50. Ebstein, F., Textoris-Taube, K., Keller, C., Golnik, R., Vigneron, N., Van den Eynde, B. J., Schuler-Thurner, B., Schadendorf, D., Lorenz, F. K., Uckert, W., Urban, S., Lehmann, A., Albrecht-Koepke, N., Janek, K., Henklein, P., Niewianda, A., Kloetzel,

- P. M., and Mishto, M. (2016) Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. *Sci Rep* **6**, 24032
51. Ebner, F., Hepworth, M. R., Rausch, S., Janek, K., Niewianda, A., Kuhl, A., Henklein, P., Lucius, R., Hamelmann, E., and Hartmann, S. (2014) Therapeutic potential of larval excretory/secretory proteins of the pig whipworm *Trichuris suis* in allergic disease. *Allergy* **69**, 1489-1497
52. Liepe, J., Mishto, M., Textoris-Taube, K., Janek, K., Keller, C., Henklein, P., Kloetzel, P. M., and Zaikin, A. (2010) The 20S Proteasome Splicing Activity Discovered by SpliceMet. *PLOS Computational Biology* **6**, e1000830
53. Zhou, Y., Liepe, J., Sheng, X., Stumpf, M. P., and Barnes, C. (2011) GPU accelerated biochemical network simulation. *Bioinformatics* **27**, 874-876
54. Team, R. C. (2014) R: A Language and Environment for Statistical Computing.
55. Peters, B., Janek, K., Kuckelkorn, U., and Holzhutter, H. G. (2002) Assessment of proteasomal cleavage probabilities from kinetic analysis of time-dependent product formation. *J Mol Biol* **318**, 847-862
56. Mishto, M., Raza, M. L., de Biase, D., Ravizza, T., Vasuri, F., Martucci, M., Keller, C., Bellavista, E., Buchholz, T. J., Kloetzel, P. M., Pession, A., Vezzani, A., and Heinemann, U. (2015) The immunoproteasome beta5i subunit is a key contributor to ictogenesis in a rat model of chronic epilepsy. *Brain Behav Immun*
57. Mishto, M., Santoro, A., Bellavista, E., Sessions, R., Textoris-Taube, K., Dal Piaz, F., Carrard, G., Forti, K., Salvioli, S., Friguet, B., Kloetzel, P. M., Rivett, A. J., and Franceschi, C. (2006) A structural model of 20S immunoproteasomes: effect of LMP2 codon 60 polymorphism on expression, activity, intracellular localisation and insight into the regulatory mechanisms. *Biol Chem* **387**, 417-429
58. Textoris-Taube, K., Keller, C., Liepe, J., Henklein, P., Sidney, J., Sette, A., Kloetzel, P. M., and Mishto, M. (2015) The T210M Substitution in the HLA-a*02:01 gp100 Epitope Strongly Affects Overall Proteasomal Cleavage Site Usage and Antigen Processing. *J Biol Chem* **290**, 30417-30428
59. Gaczynska, M., Rock, K. L., and Goldberg, A. L. (1993) Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes. *Nature* **365**, 264-267
60. Toes, R. E., Nussbaum, A. K., Degermann, S., Schirle, M., Emmerich, N. P., Kraft, M., Laplace, C., Zwinderman, A., Dick, T. P., Muller, J., Schonfisch, B., Schmid, C., Fehling, H. J., Stevanovic, S., Rammensee, H. G., and Schild, H. (2001) Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J Exp Med* **194**, 1-12
61. Gaczynska, M., Rock, K. L., Spies, T., and Goldberg, A. L. (1994) Peptidase activities of proteasomes are differentially regulated by the major histocompatibility complex-encoded genes for LMP2 and LMP7. *Proc Natl Acad Sci U S A* **91**, 9213-9217

peptide-bond hydrolysis	
k_p	peptide-bond hydrolysis rate at active site(s)
K_{aS}, K_{aP}	dissociation constant of substrate (S) and product (P) to active site(s)
n_a	Hill coefficient for binding to active site(s)
K_{iS} and K_{iP}	dissociation constant of substrate (S) and product (P) to inhibitor site(s)
n_i	Hill coefficient for binding to inhibitor site(s)
α	factor, by which K_{aS} , K_{aP} , K_{iS} and K_{iP} are multiplied
β	factor, by which k_p is multiplied upon binding to inhibitory site(s)
Transport	
k_{on}	association rate to the gate
k_{off}	dissociation rate to from gate
v_{in}	peptide influx rate
τ	peptide translocation rate inside the chamber
v_{out}	peptide efflux rate
C	capacity (maximum number of molecules inside the chamber)
transport regulation	
R_{on}	binding rate to the enhancing regulator site(s)
R_{off}	unbinding rate to the enhancing regulator site(s)
X_{enh}	strength of enhancing regulator site(s)
I_{on}	binding rate to the inhibiting regulator site(s) outside the chamber
I_{off}	unbinding rate to the inhibiting regulator site(s) outside the chamber
H	coefficient for binding to inhibiting regulator site(s) outside the chamber
Y_{in}	strength of inhibiting regulator site(s)

Table 1. List of mathematical model parameters. The table and the meaning of the parameters have been previously published by Liepe *et al.* (11) and they correspond to the parameters indicated in **Fig. 1A**.

A



B

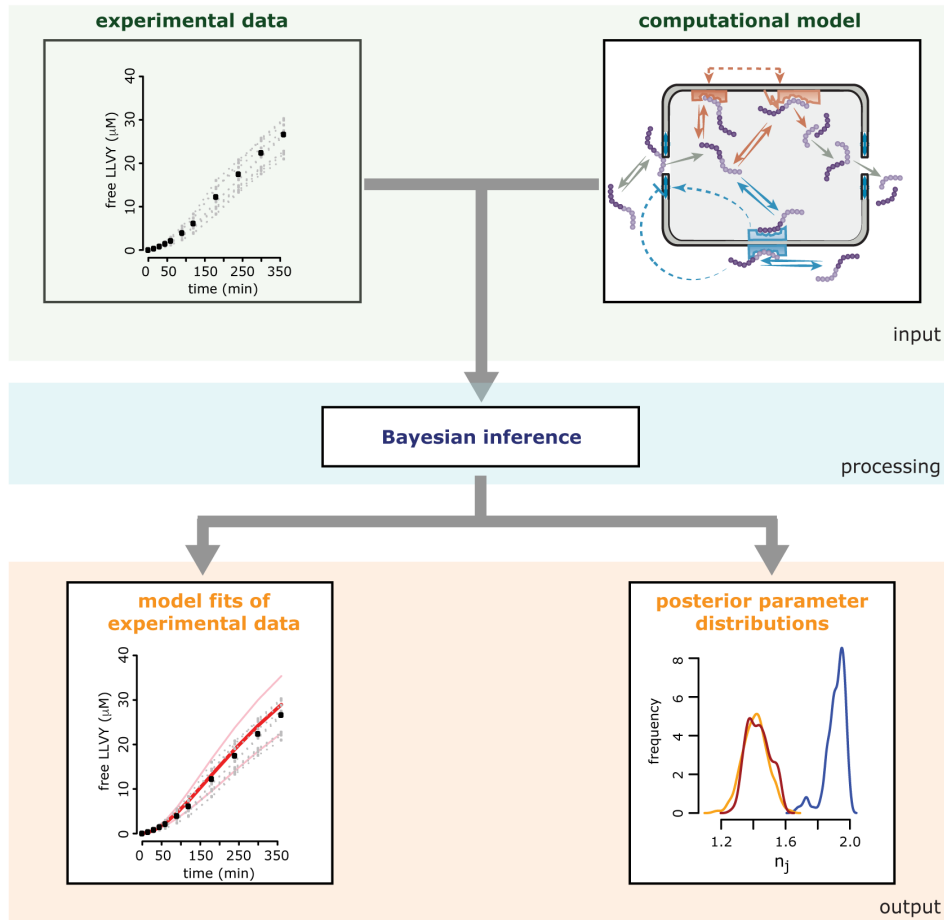


Figure 1. Overview of the computational modeling approach describing the proteasome proteolytic dynamics. In (A) the schematic of the compartmentalized proteasome model proposed by Liepe and colleagues (11) is shown. The model describes all relevant steps involved in substrate degradation. These include (i) peptide transport steps (peptide binding close to the outer site of the gate, peptide influx into the chamber, peptide translocation inside the chamber and peptide efflux out of the chamber), (ii) substrate hydrolysis steps (peptide binding to the active site and subsequent hydrolysis, and peptide binding to the non-catalytic inhibitor site) and (iii) transport regulation (peptide binding to the non-catalytic enhancer site, peptide binding to the non-catalytic inhibitor site and resulting effects on the conformation of the proteasome gate). The grey chamber represents a simplification of the 20S proteasome catalytic chamber with openings to the outside. The substrate and product peptides (purple) can enter the 20S proteasome chamber upon binding to the outer face of the gate, interact with the regulatory and catalytic sites inside the chamber, and can leave the proteasome chamber upon translocation to the proximity of the inner face of the gate. Grey arrows indicate the

transport of substrate and product peptides. The orange arrow denotes the hydrolysis reaction where a substrate peptide is transformed into product peptide, thereby releasing the fluorophore. Enhancing regulatory sites inside the chamber are shown in blue with the dashed arrows indicating their effect (transport enhancing gate conformation). The inhibiting regulatory site outside the chamber is shown in light blue with the dashed arrow indicating its effect (transport inhibiting gate conformation). The catalytic site consists of an active site (light orange) and an inactive modifier site (dark orange). For details of the model equations, model setup and model parameters, please refer to (11). In **(B)** the schematic of Bayesian inference is sketched. Computational models describing biological systems are often parameterized. These parameters can be abstract, or as it is the case here, they can be kinetics parameters with a direct physical translation. In order to learn anything from the model, it is necessary to calibrate the model against experimental data. This can be done in many fashions, however, in the last decade Bayesian inference techniques proved to be powerful model calibration tools. One of the advantages of Bayesian inference is that it estimates not only the model parameters, but also their uncertainty. In general, experimental data are collected and a computational model is formulated. Both are then used as input for the Bayesian inference algorithm (here approximate Bayesian computation). The basic concept of the algorithm is to test all sort of combinations of parameters (through a defined sampling scheme) and simulate the model with those parameter combinations. If the model simulations correspond well to the experimental data, the corresponding parameter combination is accepted; otherwise the parameter combination is discarded. This is done repeatedly until a certain number of accepted parameter combinations is reached, which then construct the so-called posterior parameter distribution. This posterior distribution contains all information about the separate model parameters as well as their dependencies among each other. The outputs of Bayesian inference are therefore the model fits of the experimental data and the posterior parameter distributions. Calibrating the same model to experimental data generated under different conditions (here different proteasome isoforms) allows us to compare the obtained posterior parameter distributions and detect which model parameters differ for the different conditions and which parameters are not influenced.

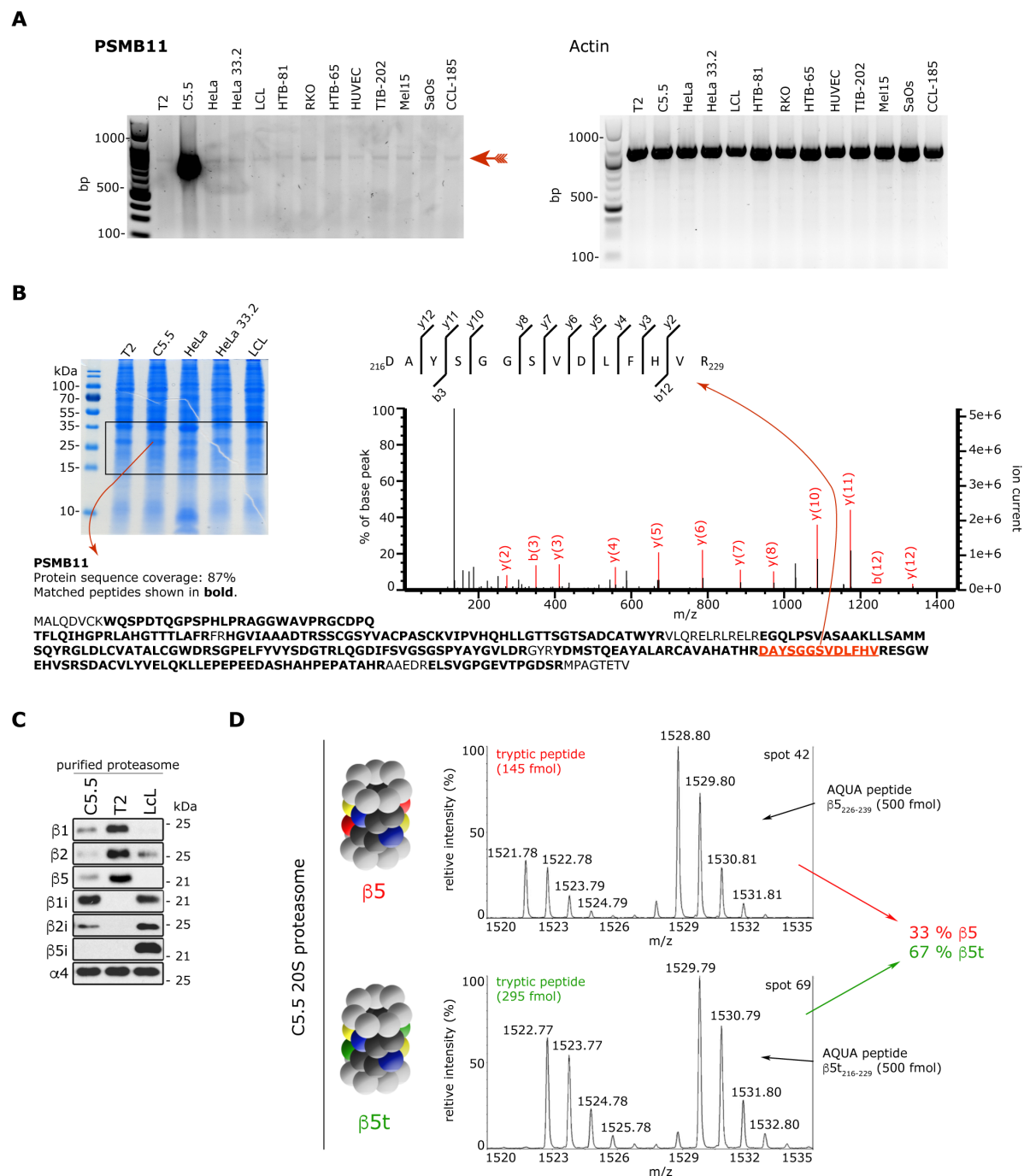


Figure 2. Catalytic β subunit expression in different human cell lines. **A)** RT-PCR products specific for PSMB11 (β 5t subunit) or actin (as control) obtained from different immortalized or tumor-derived human cell lines. The specific band of the PSMB11 is marked with a red arrow. Specificity has been confirmed by cloning and sequencing the cDNA extracted from the band. **B)** MS identification of the β 5t subunit in the cell lysate of the same cell lines showed in (shown in **A**). Cell lysates have been separated on a SDS gel and stained with Coomassie (left panel; only some representative cell samples are shown), the proteins framed in the picture have been cut, digested by trypsin and the β 5t subunit-specific peptides have been detected by MS. Only for the C5.5 cell line we have identified the β 5t subunit products (marked with bold letters: 19 specific peptides depicted below the gel) with significant MS/MS spectra (the spectrum of one of them is depicted in the right panel). **C)** Western blot assays for proteasome catalytic subunits, which have been carried out after separation of 0.5 μ g purified 20S proteasomes in SDS-Page, are shown. **D)** Relative quantification of the subunit β 5 (PSB5_human) and β 5t (PSB11_human) of purified 20S proteasome after absolute quantification with AQUA peptides. Shown are representative MALDI-MS/MS mass spectrometry spectra of the tryptic peptide

²²⁶DAYS GGAVNLYHVR₂₃₉ with $[M+H]^+_{\text{exp}} = 1521.78$ and its heavy analogue with $[M+H]^+_{\text{exp}} = 1528.80$ of $\beta 5$ subunit at spot 42 and ²¹⁶DAYS GGSVDLFHVR₂₂₉ with $[M+H]^+_{\text{exp}} = 1522.77$ and its heavy analogue with $[M+H]^+_{\text{exp}} = 1529.79$ of $\beta 5t$ subunit at spot 69. The calculation of the absolute amount of the tryptic peptides is performed by comparison of the MS peak areas with those of the corresponding AQUA peptides.

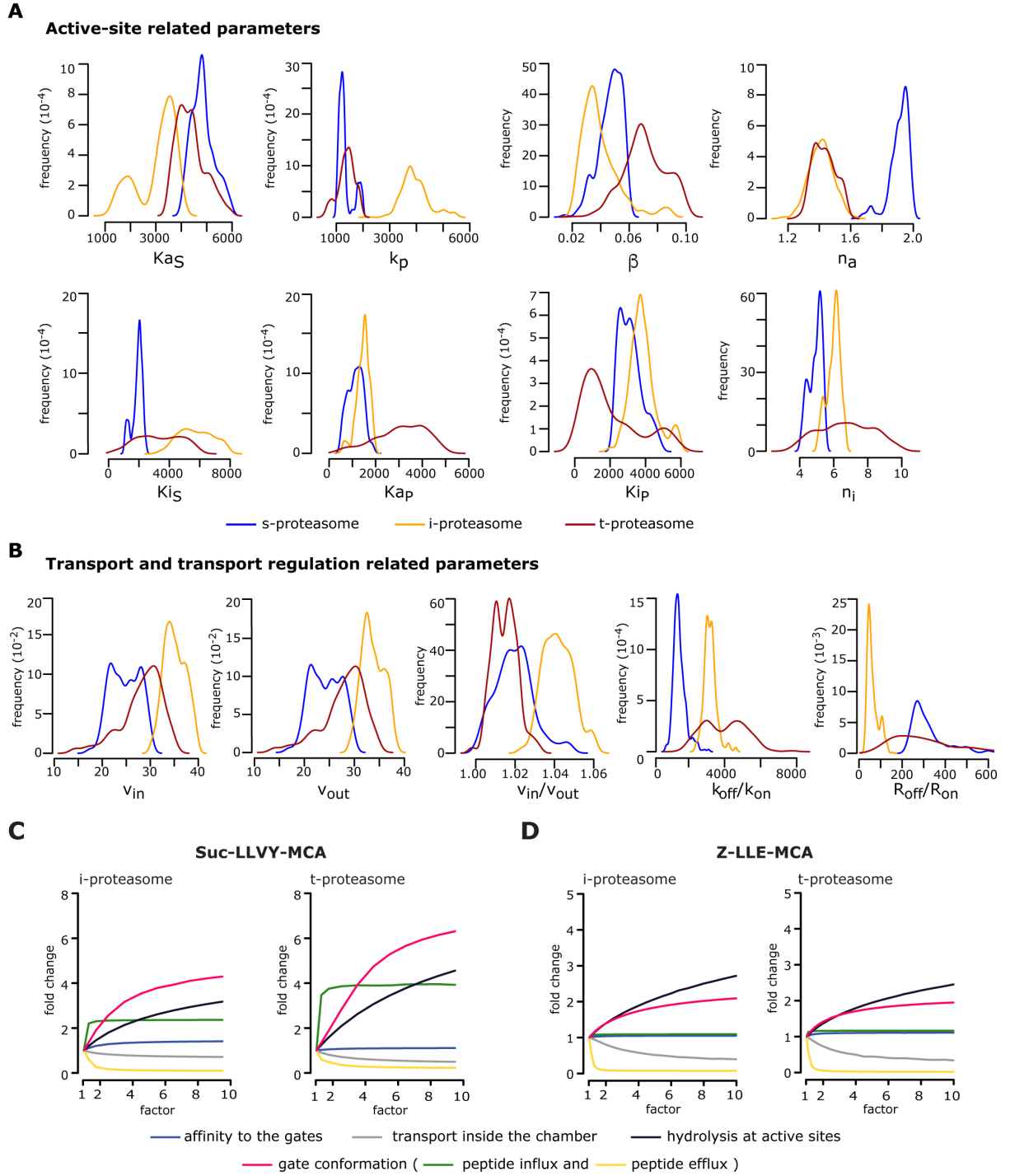


Figure 3. Computational model parameters and rate-limiting steps of proteasome isoform dynamics. **A-B)** Marginal posterior parameter distributions obtained by calibrating the proteasome kinetics model against experimental data ($n=3-5$) derived from the degradation of the substrate Suc-LLVY-MCA by 20S s-, i- and t-proteasomes. Parameters are grouped into active site-related parameters (**A**) and transport- and transport regulation-related parameters (**B**). Briefly, K_{aS} , K_{aP} are the dissociation constant of substrate (S) and product (P) to active site(s); k_p is the peptide-bond hydrolysis rate at active site(s); β is factor by which k_p is multiplied upon binding to inhibitory site(s); n_a and n_i are the Hill coefficients for binding to the active site(s) and the inhibitor site(s); K_{iS} and K_{iP} are the dissociation constants of substrate (S) and product (P) to inhibitor site(s); v_{in} and v_{out} are the peptide influx and efflux rates and v_{in}/v_{out} is their ratio; K_{off}/K_{on} is the ratio between the dissociation and the association rates to the gate; R_{off}/R_{on} is the ratio between the unbinding and binding rate to the enhancing regulator site(s). The meaning of all parameters is depicted in **Fig. 1** and **Table 1**. **C-D)** Analysis of rate-limiting steps in 20S

i- and t-proteasomes. Depicted is the fold change of product formation (y axis) upon increase (by a factor of; x axis) of a specific reaction step for the degradation of Suc-LLVY-MCA (**C**) and Z-LLE-MCA (**D**), as simulated by our computational model of the 20S proteasome dynamics. The initial substrate concentration for this analysis is 160 μ M and the fold change is determined after 60 min reaction relative to the experimentally measured proteasome kinetics (factor = 1). The mean of 1000 *in silico* predictions (colored lines) is plotted over time for the degradation of the substrates with the same initial substrate concentrations as in the experiments. The rate-limiting steps are those which the increase of leads to the largest fold change.

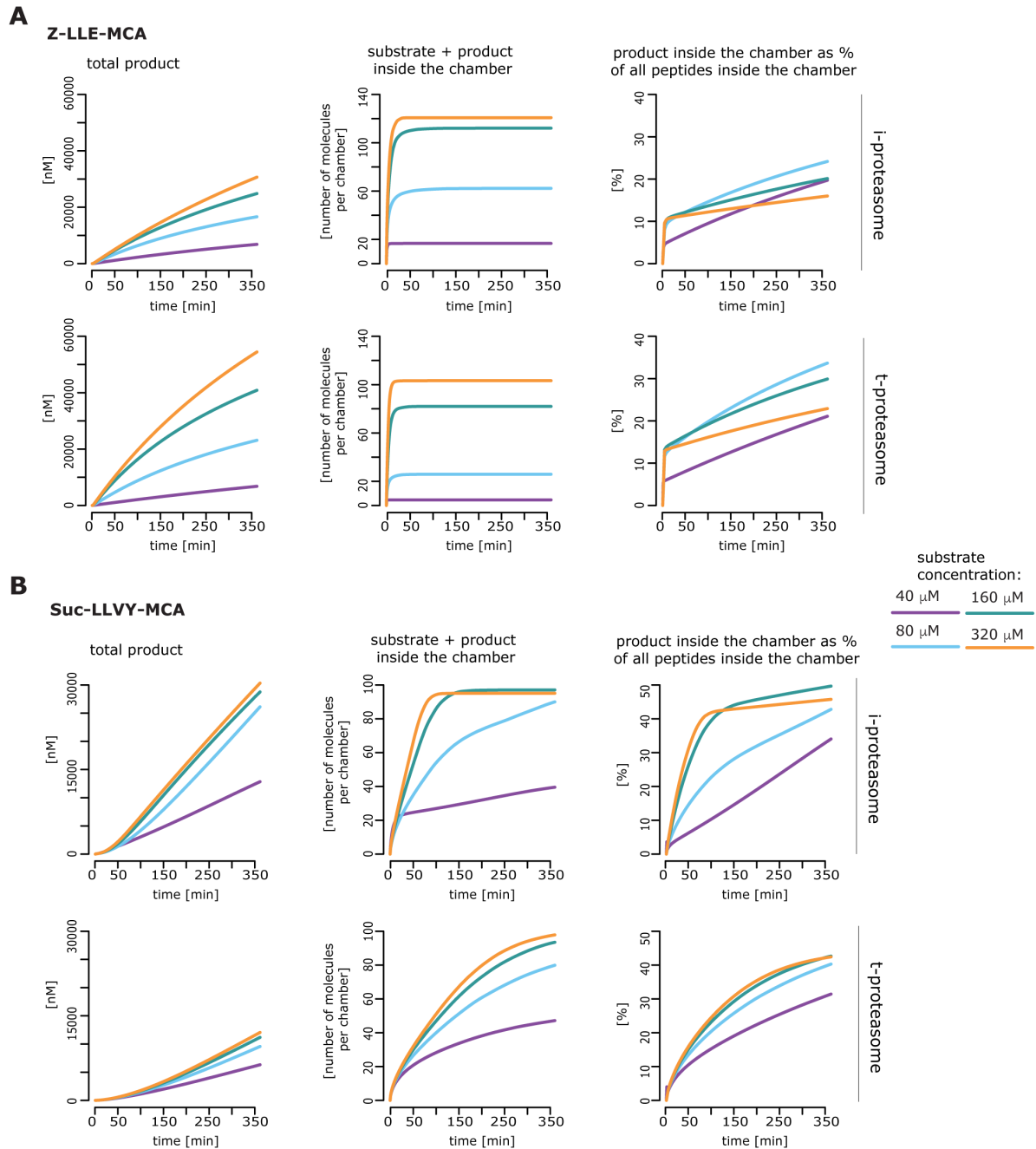


Figure 4. Simulation of the substrate and product dynamics inside the chambers of the human s-, i- and t-proteasomes. The mean of 1000 *in silico* predictions (colored lines) is plotted over time for the degradation of the substrates Z-LLE-MCA (**A**) or Suc-LLVY-MCA (**B**) with the same initial substrate concentrations as used in the independent experiments (n=3-5; see Material & Methods). The simulation has been performed for 20S i- and t-proteasomes. The number of peptide molecules (product and substrate) and the relative amount of products vs total amount of peptides inside the chambers over time has been computed by the computational model and it is based on the estimated posterior parameter distributions for each substrate and proteasome isoform.

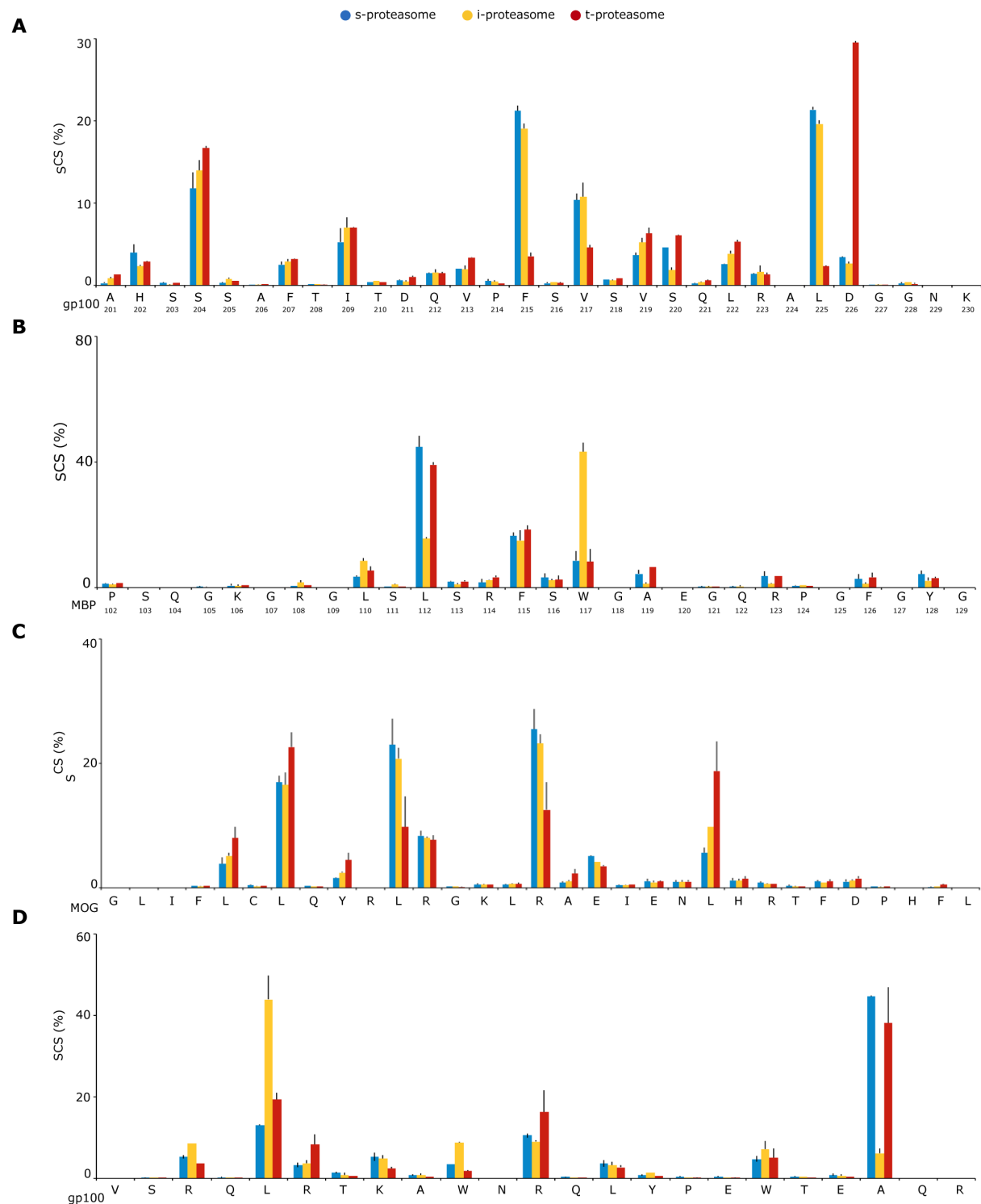


Figure 5. Different proteasome isoforms preferentially use different substrate cleavage-sites. A-D) The relative frequency of the substrate cleavage-site usage (*i.e.* the SCS of the synthetic substrates gp100₂₀₁₋₂₃₀ (A), MBP₁₀₂₋₁₂₉ (B), MOG₁₇₂₋₂₀₂ (C), and gp100₃₅₋₅₃ (D) by 20S s-, i- and t-proteasome are shown as mean of 2-4 independent experiments (bars are the SD) measured 2-3 times. Quantitative analyses are done by applying QME to the MS measurements.

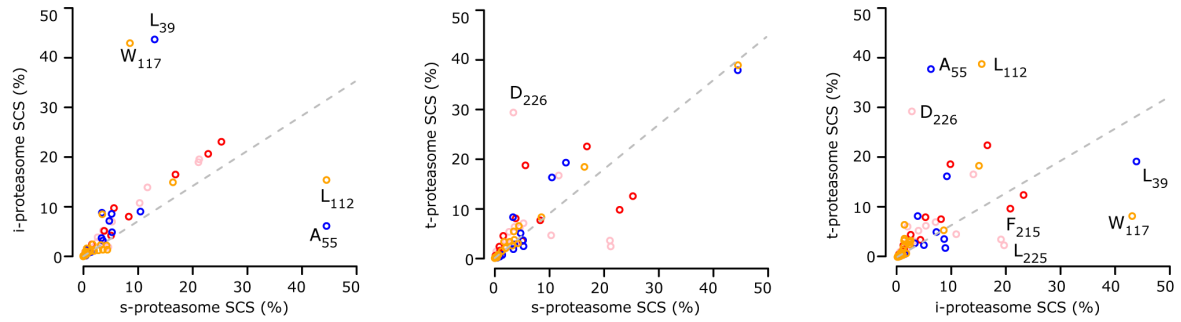


Figure 6. Proteasome isoforms overall share a common substrate cleavage-sites usage, with specific differences. The correlation between the SCSs of the synthetic substrates gp100₂₀₁₋₂₃₀ (dots in pink), gp100₃₅₋₅₃ (dots in blue), MBP₁₀₂₋₁₂₉ (dots in green), and MOG₁₇₂₋₂₀₂ (dots in black) generated in *in vitro* digestion kinetics by human 20S s-, i- or t-proteasomes. The SCSs are compared in proteasomes' pairwise. The computed Pearson's correlation coefficients and the tests for significance of these correlations are reported. Non-significant correlations are marked in bold, which depict SCSs that differ between the pair of compared 20S proteasome types. The substrate cleavage-sites which the use of is more remarkably varied between 20S proteasome isoforms are labeled.

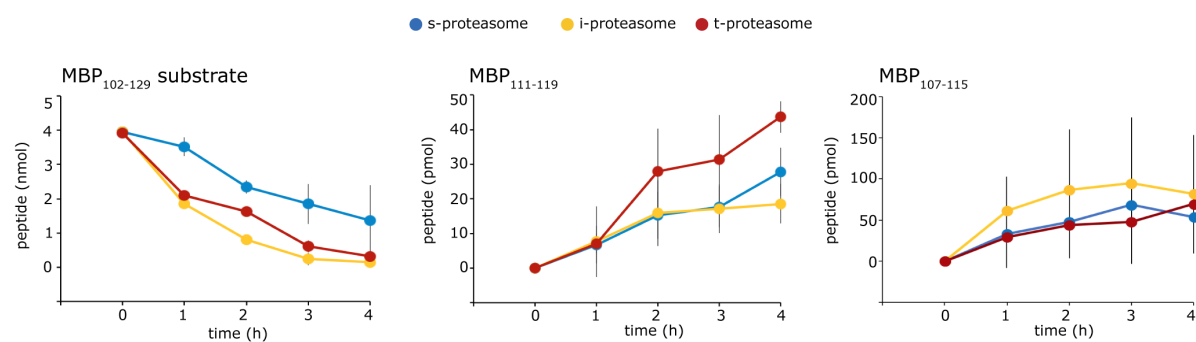


Figure 7. Proteasome isoforms generate MBP epitopes with different kinetics. The degradation kinetics of the synthetic substrates MBP₁₀₂₋₁₂₉ by 20S s-, i- and t-proteasomes, and the generation kinetics of the epitopes MBP₁₁₁₋₁₁₉ and MBP₁₀₇₋₁₁₅, is shown. The mean of 2-4 independent experiments (bars are the SD) measured in duplicate is shown. Quantitative analyses are done by applying QME to the mass spectrometry measurements.